Report on the Quality of the Voter Registration Database in the State of Indiana, Relative to Other States' Databases

Eitan D. Hersh
Assistant Professor
Department of Political Science
Yale University
New Haven, CT

September 23, 2013

Exhibit
7

I. Statement of Inquiry

1. I have been asked to research the quality of Indiana's voter registration records, relative to the quality of other states' voter registration records. In a research project that began in 2010, I, along with a colleague, designed several state-by-state measures of registration list quality. We calculated values of these measures for each state. In this report, I will analyze how Indiana fares on these measures relative to other states.

II. Background and Qualifications

2. I am an assistant professor of political science and a resident fellow of the Institution for Social and Policy Studies at Yale University in New Haven, Connecticut. I joined the faculty at Yale in 2011, in the year I received my PhD in government from Harvard University.

3. My scholarly research focuses on the topic of U.S. elections, specifically on voter behavior, election administration, and political campaigns. My methodological focus is in using new sources of individual-level, behavioral data, such as digitized voter registration records, to study politics. I have published five peer-reviewed, academic articles on these topics, in the journals *Journal of Politics*, *Quarterly Journal of Political Science, Political Analysis,* and *Political Behavior.* Currently, I am completing a book manuscript, which is under revision for publication. My other

work under review can be found in my curriculum vita, which is included at the end of this report.

4. I have worked as a consultant to the U. S. Department of Justice, Civil Rights Division, on the case *Texas v. Holder*. I continue to work as a consultant to the Department of Justice.

5. I am compensated for my work at a rate of $250 per hour.

III. Method of Study

6. In 2010, I began a research project with Stephen Ansolabehere, Professor of Government at Harvard, to generate measures of the quality of voter registration lists nationwide. We first released our report detailing our method and initial results in July 2010, in a working paper titled "The Quality of Voter Registration Records: A State-by-State Analysis." The paper was circulated as a Caltech/MIT Voting Technology Project report. We have since revised the paper, and it is being published as a chapter in a peer-reviewed edited volume, titled *The Measure of American Elections*, edited by Barry C. Burden and Charles Stewart III, forthcoming with Cambridge University Press. The method and analysis summarized here is derivative of work in the forthcoming book chapter.

7. My work with Professor Ansolabehere on measuring registration quality came out of an understanding that the national political parties, as well as private data vendors that contract with political campaigns and parties, have compiled voter registration data from all states. Voter registration is decentralized in the United States; states and counties manage the voter lists within their jurisdictions. However, because national campaigns and parties use voter registration data to contact voters, they have put resources into compiling and cleaning voter lists for all jurisdictions. Consequently, the political parties and their vendors maintain national voter files. By contracting with a political data vendor that maintains a nationwide database of registered voters, Professor Ansolabehere and I could estimate comparable measures of the quality of registration records across jurisdictions.

8. The vendor we contracted with is called Catalist, LLC, a Washington, DC-based firm that vends data to Democratic campaigns and progressive organizations. Catalist also contracts with academic institutions, selling data used in scholarly research. Catalist builds its national voter file is by acquiring official voter registration lists from state and county election offices. It then cleans the records by performing operations such as identifying duplicative records, records of people thought to be deceased, records of individuals who have changed residences, and records of individuals that are missing key information such address and birthdate fields. Catalist performs these kinds of data-cleansing tasks so that its clients do not waste resources, for example by sending canvassers and mailers to addresses of people who have moved or to homes of deceased voters. Indeed, one reason why a political campaign might contract with

Catalist rather than just acquiring a voter file directly from an election authority is because of Catalist's efforts to clean lists of obsolete and inaccurate records. The analysis that Professor Ansolabehere and I have done using Catalist's records leverages the information Catalist has generated by cleaning records, such as when Catalist finds incorrect address information or records of voters presumed to be deceased.

9. In simple terms, Catalist receives registration files from the state and cleans those files. In cleaning the files, it flags records that appear to be problematic for a variety of reasons. Professor Ansolabehere and I used Catalist's cleaning process to study the quality of each state's records. The primary measures of list quality we developed and calculated were as follows:

   a. Missing Address information. Catalist flags records that are missing a city, state, or ZIP code field in their mailing address.

   b. Undeliverable Addresses. Catalist links the addresses from voter registration files to the U.S. Post Office's Coding Accuracy Support System (CASS). This system estimates whether an address is a mailable address.

   c. Birthday distribution. We asked Catalist to calculate the distribution of birthdays (month and day) for all voters in each state. If voters' birthdays are recorded accurately, there should be a similar rate of birthdays falling on each day of the year. If birthdays are systematically recorded incorrectly, there will deviations from the expected distribution.

d. Registration Date Coverage. Catalist records whether or not a registered voter is listed with a registration date.

e. Registration Distribution. We calculated the proportion of registrants listed with a registration date as January 1. January 1 is unlikely to be the actual date of registration for most registrants, but some states are in the practice of using January 1 as a default registration date.

f. Deceased voters. Catalist matches its records to both the Social Security Administration's Death Index and to commercial databases that estimate whether a person is deceased.

g. Deadwood. Catalist estimates whether each record is obsolete or not. Obsolete records are often referred to as ``deadwood.'' Catatlist's ``deadwood'' model takes into account a voter's listed birth year, an estimate of whether the voter is deceased, inactive registration status, evidence that the voter has moved and re-registered elsewhere, and the extent to which the voter has participated in prior elections.

h. Vote History Discrepancy. Voter registration lists contain information about which voters participated in each election. The overall turnout tally can thus be estimated by the number of registrants marked as having voted. Turnout can alternatively be measured by the number of ballots cast, as reported by each state or county. These two measures of turnout should be similar. We measure the discrepancy between the two measures in each state. To measure the discrepancy, we first measure the deviation between ballots cast (as reported by election officials) and turnout as reported in the voter file by

county within a state. We take the absolute value of the counties' deviations and sum them. We then divide the total absolute county-by-county deviations by the total number of votes cast in the state.

10. The measures we estimate bear on several governmental functions of voter registration lists. First, in order for election offices to communicate with voters (for example, to notify them of upcoming elections or to determine whether they are still valid voters), they must maintain accurate mailing address information. Records that are missing address information or that contain undeliverable addresses are thus problematic. Second, accurate birthdate information may be necessary to validate voters at the polls in certain jurisdictions. Third, as part of the process to audit election results, it could be important for states to have an accurate record of the number of registered voters listed as having casted votes. Finally, to maintain clean voter registration lists, it is important for states to keep track of voting records that are obsolete. To remove voters from registration lists in accordance with the National Voter Registration Act, states must maintain accurate information about when a voter entered the registration system and whether the voter participated in recent elections, as well as accurate mailing information in order to contact the voter. Relatedly, estimates of deadwood on voter files may help gauge the extent to which states are cleaning voter files of obsolete records.

11. States and counties update their voter registration systems dynamically. Catalist acquires updated lists from states several times a year. The data in this analysis were

collected from Catalist's records between March and June 2010. The evidence here reflects the snapshot of the voter files available at that time.

IV. Analysis

12. In all states, the number of registered voters identified in the Catalist voter file totaled 185,445,103. This number is calculated after Catalist collapsed duplicative records. It was not possible to accurately gauge the number of duplicative records that Catalist collapsed in each state. In the state of Indiana, Catalist's records indicated 4,206,831 registered voters. Of these, 106,081 were listed as inactive.

13. Table 1 shows several statistics for seven of the eight measures of registration list quality described above. For each measure, the table reports the mean value in the states, the value for the median state, the number of states included in the analysis, and the value for the state of Indiana.

14. The first row of data refers to the number of registration records in the state that Catalist identified as having a missing city, state, or ZIP code field. In the 50 states, on average, only 0.2% of records (i.e. 1 in 500 records) had a missing city, state, or ZIP code field. In the typical, or median, state, only 0.1% of records contained a missing field of this kind. On this measure, Indiana has a lower percentage of records with missing address fields than the average and the same percentage as the median state.

15. The second row of data reflects on the percentage of records containing mailing addresses thought to be undeliverable or probably undeliverable. This designation is determined by Catalist linking its records to the Post Office's CASS system. On average, 4.3% of records are estimated as having undeliverable or probably undeliverable addresses. In Indiana, 4.1% of records are estimated as having undeliverable or probably undeliverable addresses. This value is lower than the average and the same as the median state.

16. The third row of data reflects on the percentage of records that Catalist finds to have a missing date of registration. On average, 2.9% of records were found to have a missing registration date. In the median state, only 0.2% of records were found to have a missing registration data. The difference between the median and mean value is attributable to the fact that a few states had very large numbers of records with missing registration dates (i.e. more than 10%), whereas the majority of states had very small number of records with missing registration date. Indiana's estimated rate of records with missing registration date is below the mean and consistent with the typical state.

17. The fourth row of data reflects on whether states had an unusual distribution of registration dates. In particular, we calculated the percentage of records that were listed with January 1 registration dates. On average, 2.9% of registration records had a registration date of January 1. In the median state, 0.4% of records had a January 1 registration date. In Indiana, 0.9% of records had a January 1 registration date. This value is lower than the average value, but slightly higher than the median state.

18. The fifth row of data reflects on Catalist's estimate of the number of records thought to be associated with deceased voters. On average, 0.9% of records are estimated to be associated with deceased voters. In the median state, 0.8% of records are estimated to be associated with deceased voters. In Indiana, 0.7% of records are estimated with be associated with deceased voters.

19. The sixth row of data reflects on Catalist's estimate of records that are likely or probably deadwood. In other words, these are records that Catalist flags as likely or probably obsolete, based on Catalist's statistical model. On average, 3.9% of records are estimated as likely or probably deadwood. In the median state, 4.0% of records are estimated to be likely or probably deadwood. In Indiana, only 2.5% of registration records are estimated to be likely or probably deadwood.

20. The final two rows of data in Table 1 reflect on the discrepancy in turnout (in 2008 and in 2006) between a registration-based estimate and a ballots-cast-based estimate. A value of 2.5% on the 2008 vote discrepancy measure can be interpreted as the percent of votes cast that were inconsistent with the record of turnout as shown on the voter file. On the 2008 measure, Indiana's value is between the median and the mean. In the 2006 measure, Indiana's value is below the median and the mean.

21. The one measure defined above not listed in Table 1 is represented in Figure 1. We calculated the percentage of records in each state with birthdays falling on each day of the year. In Figure 1, each day of the year is represented on the horizontal axis. The percentage of registrants with birthdays on each day is represented on the vertical

axis. If birthdays are recorded in voter files accurately, there should be a similar rate of birthdays falling on each day of the year. We were able to ascertain the distribution of birthdays in 35 states, Indiana among them. Figure 1 shows the birthdays in 29 states with the expected distribution and three of the six states that had unusual distributions. As the Figure shows, in New Hampshire, Wyoming, and Mississippi, there is evidence that many registrants are listed with birthdays on the first day of the month. This is reflected in the lower average lines for these states and the sharp periodic spikes. In Indiana (represented by a dotted black line), like the 28 states represented in light gray, there is a fairly uniform rate of birthdays falling on each day of the year.

V. Conclusion

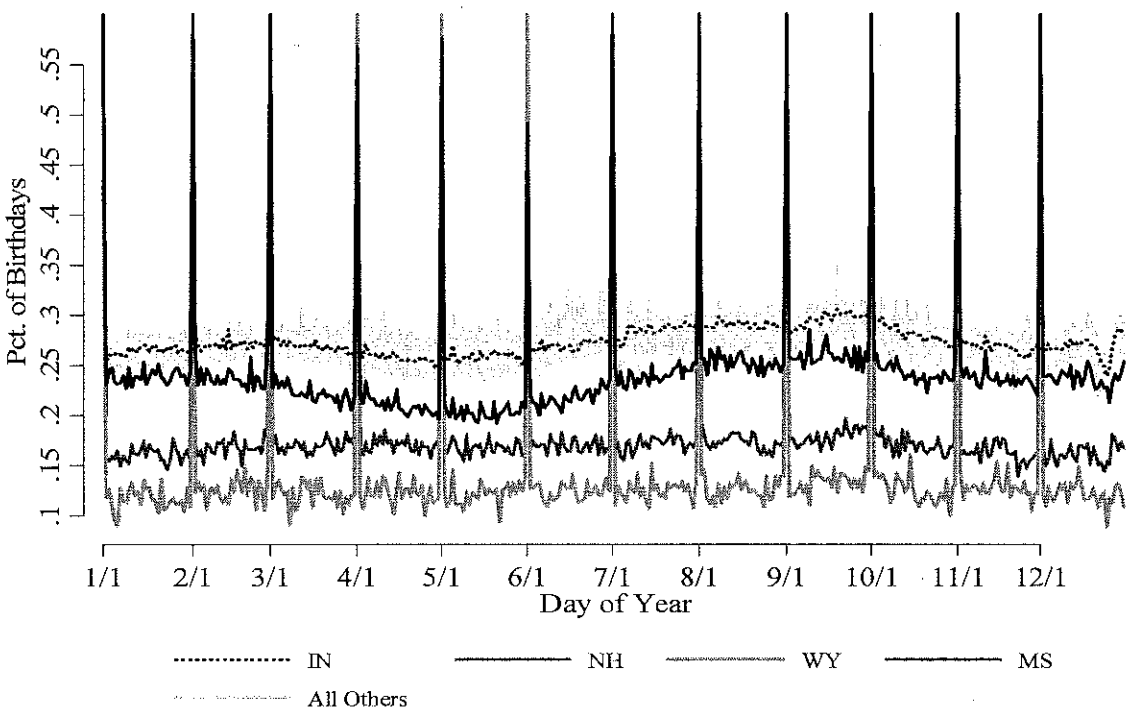
22. On each of the eight measures estimated, Indiana's voter registration data appears to be typical of registration data across the country. The rate of missing address fields, undeliverable addresses, missing and January 1 registration dates, records of voters presumed to be deceased, deadwood, vote history discrepancies, and unusual birthdate data, are all lower in the state of Indiana than is the average across states. Indiana appears to have similar - and on most measures better - records than the median state. On none of the measures does Indiana appear to be an outlier with records that appear markedly different or markedly more problematic than the national norm.

Table 1. Results for Indiana in Comparison with Other States on Seven Measures of Registration List Quality.

	Mean State	Median State	Observations	Indiana
Missing Address Fields	0.2%	0.1%	50	0.1%
Undeliverable/ Probably Undeliverable Address	4.3%	4.1%	50	4.1%
Missing Registration Date	2.9%	0.2%	50	0.2%
January 1 Registration Date	1.6%	0.4%	50	0.9%
Deceased Voters	0.9%	0.8%	50	0.7%
Likely or Probably Deadwood	3.9%	4.0%	50	2.5%
Vote History Discrepancy from Official Tally (2008)	2.5%	1.5%	48	2.0%
Vote History Discrepancy from Official Tally (2006)	4.4%	2.8%	47	1.3%

Note: Data generated from research reported in Stephen Ansolabehere and Eitan Hersh. Forthcoming. "Voter Registration: The Process and Quality of Lists." *The Measure of American Elections*. Eds. Barry Burden and Charles Stewart III. Cambridge University Press.

Figure 1. Distribution of Birthdays in Indiana in comparison with Three States with Unusual Distributions and 28 States with Typical Distributions.



Note: Data generated from research reported in Stephen Ansolabehere and Eitan Hersh. Forthcoming. "Voter Registration: The Process and Quality of Lists." *The Measure of American Elections*. Eds. Barry Burden and Charles Stewart III. Cambridge University Press.

Eitan D. Hersh

Department of Political Science
Yale University
77 Prospect Street, Room A104
New Haven, CT 06511
Tel: 203-436-9061
Email: eitan.hersh@yale.edu
Web: www.eitanhersh.com

EMPLOYMENT

Assistant Professor
Yale University, 2011-Present

EDUCATION

Ph.D. in Political Science
Harvard University, 2011

M.A. in Political Science
Harvard University, 2010

B.A. in Philosophy
Tufts University, 2005

BOOK MANUSCRIPT

Hacking the Electorate: Mass Mobilization At the Mercy of Data. Book Manuscript under revision for Cambridge University Press.

Book Conference held 08.28.13. Participants: Barry Burden, Robert Erikson, Rick Hasen, Sunshine Hillygus, Brian Schaffner, John Sides, and Lynn Vavreck

ARTICLES

Eitan Hersh and Brian F. Schaffner. 2013. "Targeted Campaign Appeals and the Value of Ambiguity." *Journal of Politics.* 75 (2): 520-534.

Stephen Ansolabehere and Eitan Hersh. 2012. "Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate." *Political Analysis* 20(4): 437-459.

Stephen Ansolabehere, Eitan Hersh, and Kenneth Shepsle. 2012. "Movers, Stayers, and Registration: Why Age is Correlated with Registration in the U.S." *Quarterly Journal of Political Science* 7(4): 333-363.

Bernard Fraga and Eitan Hersh. 2011. "Voting Costs and Voter Turnout in Competitive Elections." *Quarterly Journal of Political Science* 5(4): 339-356.

Eitan Hersh. 2011. "Primary Voters Versus Caucus Goers and the Peripheral Motivations of Political Participation. *Political Behavior.* 34(4): 689-718.

ARTICLES UNDER REVIEW

"The Long-Term Effect of September 11 on the Political Behavior of Victims' Families and Neighbors." *Under Review at PNAS*

"A Direct-Observation Approach to Identify Small-Area Variation in Political Behavior." Co-authored with Clayton Nall. *Under Review at QJPS*

"Elite Perceptions of Electoral Closeness: Fear in the Face of Uncertainty or Overconfidence of True Believers." Co-authored with Ryan Enos. *Under Review at AJPS.*

"Third Party Core Voters in Congressional Elections." Co-authored with Stacey Chen and Andrew Connery. *Under Review at LSQ*

WORKS IN PROGRESS

"The Information Burden of Voter ID Laws: Evidence from Google Searches." Co-authored with Benjamin Reis and Michael Young.

"Dynamic Voting in a Dynamic Campaign: Three Theories of Early Voting." Co-authored with Vivekinan Ashok, Daniel Feder, and Mary McGrath.

"Party Activists as Campaign Advertisers: Why the Ground Campaign Can't Move to the Middle." Co-authored with Ryan Enos.

"Are Close Elections Random?" Co-authored with Justin Grimmer, Brian Feinstein, and Daniel Carpenter.

OTHER WRITING

Stephen Ansolabehere and Eitan Hersh. 2011. "Who *Really* Votes?" In *Facing the Challenge of Democracy*, eds. Paul M. Sniderman and Benjamin Highton. Princeton: Princeton University Press.

Eitan D. Hersh, "A Caucus-Goer's Community," *Reuters*, January 3, 2012.

"The Quality of Voter Registration Records: A State-by-State Analysis," Caltech/MIT Voting Technology Project Report, July 14, 2010 (with Stephen Ansolabehere).

"Voter Registration List Quality Pilot Studies," Caltech/MIT Voting Technology Project Report, June 8, 2010 (with Stephen Ansolabehere, David Doherty, and Alan Gerber).

TEACHING

PLSC 232: "Information, Technology, and Politics" (Yale)

PLSC 229: "Election Rules and Campaign Strategy" (Yale)

PLSC 217/853: "U.S. National Elections" (Yale)

REFEREE SERVICE

American Political Science Review, *British Journal of Political Science*, *Election Law Journal*, *Electoral Studies*, *Journal of Politics*, *Legislative Studies Quarterly*, *Political Analysis*, *Political Behavior*, *Review of Economics and Statistics*, *State Politics and Policy Quarterly.*

GUEST SPEAKING ENGAGEMENTS

I have given presentations at annual meetings of the American Political Science Association, Midwest Political Science Association, and Society for Political Methodology.

I have been invited to give guest lectures at Harvard, UCLA, University of Massachusetts-Amherst, University of Michigan, Princeton, and Yale Law School.

CONSULTING

Statistical consultant, *Texas v. Holder*, U.S. D.O.J., Civil Right Division

Commissioner, Commission on Youth Voting and Civic Knowledge, CIRCLE, 2012-14